

Diminos Case Study

```
In [1]: import pandas as pd
# import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [41]: data = pd.read_csv("diminos_data.csv")
data.drop("order_id", axis=1, inplace=True)
```

```
In [42]: from datetime import datetime

data.order_placed_at = data.order_placed_at.apply(datetime.fromisoformat)
data.order_delivered_at = data.order_delivered_at.apply(datetime.fromisoformat)
data.shape
```

```
Out[42]: (15000, 2)
```

```
In [43]: data.head()
```

```
Out[43]:
```

	order_placed_at	order_delivered_at
0	2023-03-01 00:00:59	2023-03-01 00:18:07.443132
1	2023-03-01 00:03:59	2023-03-01 00:19:34.925241
2	2023-03-01 00:07:22	2023-03-01 00:22:28.291385
3	2023-03-01 00:07:47	2023-03-01 00:46:19.019399
4	2023-03-01 00:09:03	2023-03-01 00:25:13.619056

```
In [44]: data["time_taken_hrs"] = data.order_delivered_at - data.order_placed_at
data.time_taken_hrs = data.time_taken_hrs.apply(lambda x: x.total_seconds
```

```
In [45]: data.head()
```

```
Out[45]:
```

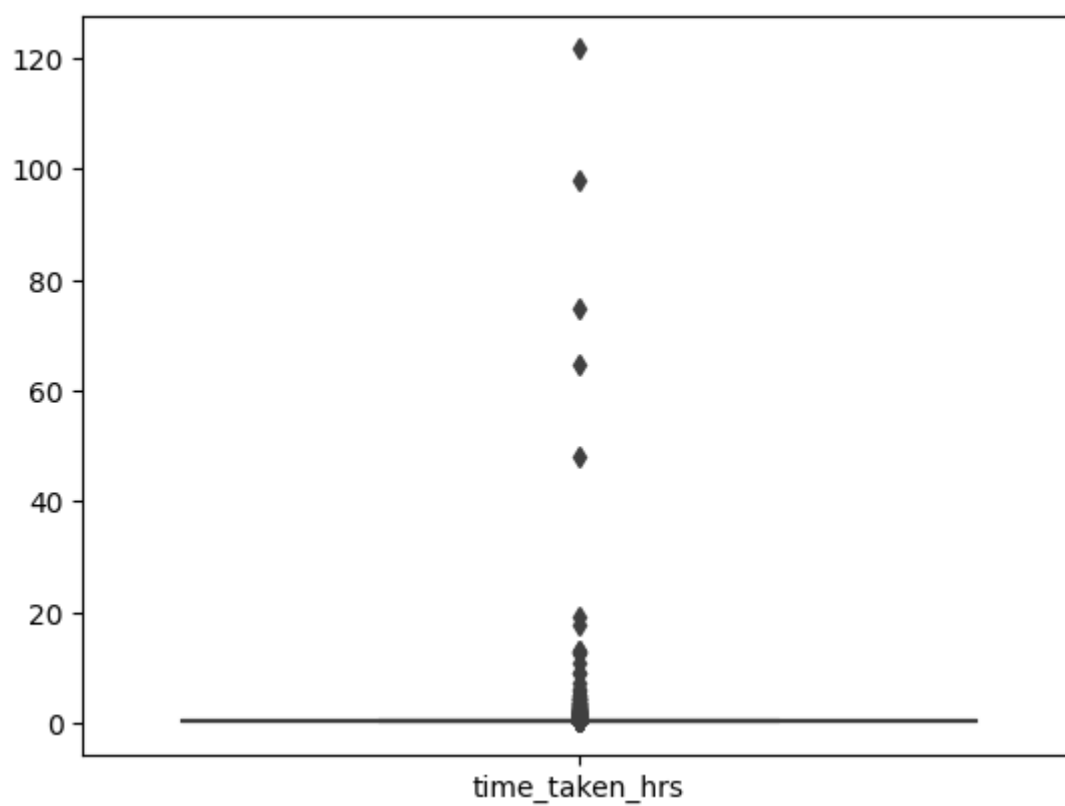
	order_placed_at	order_delivered_at	time_taken_hrs
0	2023-03-01 00:00:59	2023-03-01 00:18:07.443132	0.285679
1	2023-03-01 00:03:59	2023-03-01 00:19:34.925241	0.259979
2	2023-03-01 00:07:22	2023-03-01 00:22:28.291385	0.251748
3	2023-03-01 00:07:47	2023-03-01 00:46:19.019399	0.642228
4	2023-03-01 00:09:03	2023-03-01 00:25:13.619056	0.269616

```
In [46]: data.describe()
```

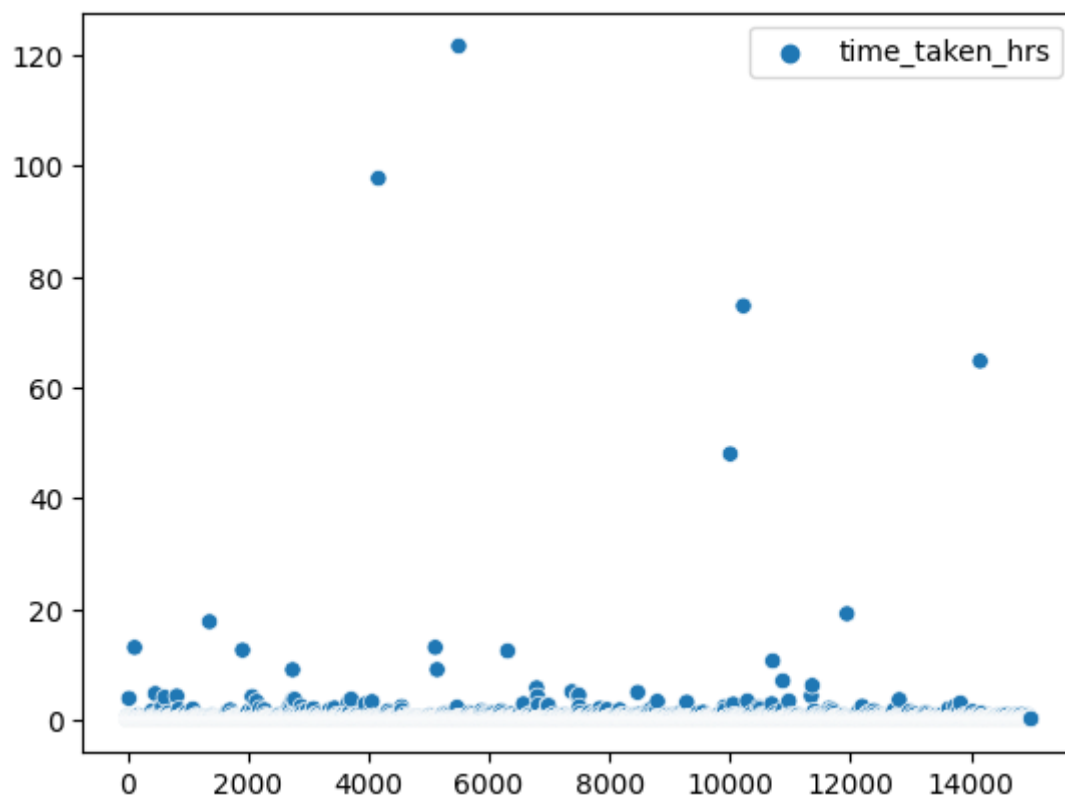
```
Out [46]:
```

	time_taken_hrs
count	15000.000000
mean	0.341656
std	1.602673
min	0.250000
25%	0.254580
50%	0.263300
75%	0.287994
max	121.663856

```
In [47]: sns.boxplot(data)  
plt.show()
```



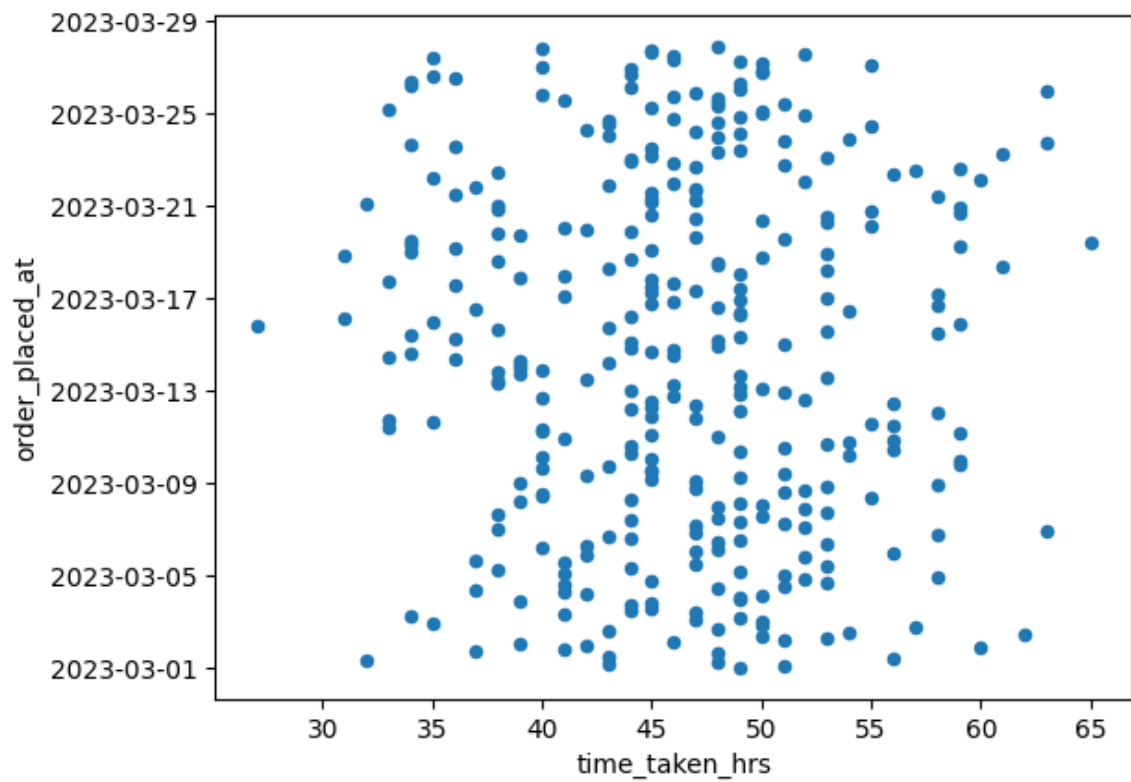
```
In [48]: sns.scatterplot(data)  
plt.show()
```



According to the above plots we have some outliers but otherwise the delivery time is always under 20 minutes

```
In [49]: data1 = data.drop("order_delivered_at", axis=1)
         grouped = data1.groupby(pd.Grouper(key="order_placed_at", freq="2H"))
         grouped = pd.DataFrame(grouped.size().reset_index(name='time_taken_hrs'))
```

```
In [50]: grouped.plot(x="time_taken_hrs", y="order_placed_at", kind="scatter")
         plt.show()
```



Almost all orders take more than 30 hours to be delivered

```
In [51]: grouped[grouped["time_taken_hrs"] < 24].shape
```

```
Out[51]: (0, 2)
```

All orders take more than a day to be delivered